

Biostat_projet

Claire et Lucie

01/12/2021

Le Danemark est un pays scandinave. Pour ce pays, nous allons étudier les différentes causes de mortalités en fonction de l'âge des individus, de la période où elles sont mortes.

Explications du jeu de données

```
mort <- read.csv("mortDK.csv", header = TRUE, sep = ";")
mort <- mort %>% rename("taux.deces" = rt,
                         "Infections" = r1,
                         "Cancer" = r2,
                         "Tumeurs.Bénignes" = r3,
                         "Endocrine" = r4,
                         "Maladies.Sang" = r5,
                         "Systeme.Nerveux" = r6,
                         "Cerebrovasculaire" = r7,
                         "Cardiaque" = r8,
                         "Maladies.Respiratoires" = r9,
                         "Maladies.Foie" = r10,
                         "Maladies.Digestives" = r11,
                         "Maladie.Génito_urinaire" = r12,
                         "Non.Definis" = r13,
                         "Mort.Naturelle" = r14,
                         "Mort.Violente" = r15) %>%
  select(-i..)
mort$sex <- mort$sex %>% factor( labels = c("H", "F"))
mort$periode <- mort$periode %>% factor( labels =c("43-47", "48-52",
                                                       "53-57", "58-62",
                                                       "63-67", "68-72",
                                                       "73-77", "78-82",
                                                       "83-87", "88-92"))
mort$age <- mort$age %>% factor()
```

Le jeux de données est composé de 1820 observations et 21 variables. Ces variables sont l'âge, le sexe, la période, le nombre de décès, le taux de décès et 15 variables sur l'identification du décès. Lorsque qu'on parle d'identification du décès, on parle si la personne est morte d'un cancer, d'une mort naturelle, par exemple.

La variable *taux.deces* est calculée d'après la formule suivante :

$$\text{taux.deces} = \frac{1000 \times \text{nb.deces}}{PA}$$

Les 15 variables d'identification du décès sont calculés de la même manière.

```
summary(mort)
```

```
##      age      periode    sex       PA      nb.deces
## 0     : 20   43-47 :182   H:910   Min.   : 2750   Min.   : 13.0
## 1     : 20   48-52 :182   F:910   1st Qu.: 89820  1st Qu.: 155.0
## 2     : 20   53-57 :182          Median :140311  Median : 545.5
## 3     : 20   58-62 :182          Mean   :126108  Mean   : 1288.2
## 4     : 20   63-67 :182          3rd Qu.:161190  3rd Qu.: 2216.2
## 5     : 20   68-72 :182          Max.   :224654  Max.   :16143.0
## (Other):1700 (Other):728
##      taux.deces      Infections      Cancer      Tumeurs.Bénignes
## Min.   : 0.1250   Min.   :0.0000   Min.   : 0.010   Min.   :0.0000
## 1st Qu.: 0.9602   1st Qu.:0.0160   1st Qu.: 0.114   1st Qu.:0.0080
## Median : 3.7620   Median :0.0855   Median : 1.089   Median :0.0345
## Mean   : 31.5585  Mean   :0.2249   Mean   : 5.303   Mean   :0.1886
## 3rd Qu.: 29.2005  3rd Qu.:0.3080   3rd Qu.: 8.064   3rd Qu.:0.2032
## Max.   :386.0400  Max.   :3.4940   Max.   :34.902   Max.   :2.8850
##
##      Endocrine      Maladies.Sang      Systeme.Nerveux      Cerebrovasculaire
## Min.   :0.0000   Min.   :0.000   Min.   : 0.000   Min.   : 0.000
## 1st Qu.:0.0140   1st Qu.:0.000   1st Qu.: 0.037   1st Qu.: 0.011
## Median :0.0710   Median :0.012   Median : 0.113   Median : 0.118
## Mean   :0.3615   Mean   :0.101   Mean   : 1.143   Mean   : 3.866
## 3rd Qu.:0.4665   3rd Qu.:0.065   3rd Qu.: 0.411   3rd Qu.: 2.533
## Max.   :3.6330   Max.   :2.299   Max.   :118.890  Max.   :52.136
##
##      Cardiaque      Maladies.Respiratoires      Maladies.Foie      Maladies.Digestives
## Min.   : 0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.: 0.0340   1st Qu.:0.0220   1st Qu.:0.0060   1st Qu.:0.0130
## Median : 0.5305   Median :0.1165   Median :0.1100   Median :0.0865
## Mean   : 13.2173  Mean   : 2.4315  Mean   :0.4175   Mean   : 0.6074
## 3rd Qu.: 10.7007  3rd Qu.: 1.5092  3rd Qu.:0.5677   3rd Qu.:0.6130
## Max.   :164.3500  Max.   :42.0310  Max.   :4.6050   Max.   : 7.5020
##
##      Maladie.Génito_urinaire      Non.Definis      Mort.Naturelle      Mort.Violente
## Min.   : 0.0000   Min.   :0.0000   Min.   : 0.0000   Min.   : 0.0220
## 1st Qu.: 0.0130   1st Qu.:0.0100   1st Qu.: 0.0290   1st Qu.: 0.3018
## Median : 0.0860   Median :0.0660   Median : 0.0625   Median : 0.7060
## Mean   : 1.0289   Mean   : 0.8154  Mean   : 0.3209   Mean   : 1.5314
## 3rd Qu.: 0.6138   3rd Qu.: 0.5503  3rd Qu.: 0.1735   3rd Qu.: 1.2528
## Max.   :23.9850   Max.   :27.2030  Max.   :50.2120   Max.   :23.4940
##
```

TODO : description du data avec stat descriptives ?

```
head(mort)
```

```
##   age periode sex      PA nb.deces taux.deces Infections Cancer Tumeurs.Bénignes
## 1   0   43-47   H 224654     12072    53.736     3.494  0.098      0.085
## 2   0   43-47   F 212730     8688    40.841     2.994  0.061      0.056
```

```

## 3 0 48-52 H 202692 7035 34.708 1.692 0.074 0.039
## 4 0 48-52 F 192602 5001 25.965 1.516 0.078 0.067
## 5 0 53-57 H 193164 5797 30.011 0.601 0.119 0.036
## 6 0 53-57 F 182400 3991 21.880 0.450 0.088 0.044
## Endocrine Maladies.Sang Système.Nerveux Cerebrovasculaire Cardiaque
## 1 0.347 0.089 1.282 0.013 0.036
## 2 0.259 0.127 1.011 0.009 0.033
## 3 0.222 0.074 0.834 0.015 0.049
## 4 0.119 0.036 0.597 0.021 0.026
## 5 0.140 0.026 0.766 0.047 0.047
## 6 0.088 0.016 0.488 0.000 0.055
## Maladies.Respiratoires Maladies.Foie Maladies.Digestives
## 1 10.826 0.027 5.092
## 2 8.739 0.038 3.685
## 3 6.276 0.054 1.919
## 4 4.875 0.042 1.231
## 5 3.018 0.031 1.066
## 6 2.489 0.033 0.691
## Maladie.Génito_urinaire Non.Definis Mort.Naturelle Mort.Violente
## 1 0.125 1.340 29.962 0.921
## 2 0.132 1.020 21.976 0.700
## 3 0.049 0.745 21.718 0.947
## 4 0.062 0.633 15.919 0.742
## 5 0.031 0.564 22.810 0.709
## 6 0.027 0.378 16.541 0.493

```

Entre les années 1943 et 1948, 224 654 nourrissons étaient de sexe masculins. 12 072 d'entre eux sont morts. Ce qui représente environ 53 individus issus de cette catégorie sur 1000 sont morts sur cette période. **TODO plus détailler une autre ligne**

Modèle à choisir, lequel on fait (beaucoup de variables, avec pleins de catégories)

Modalité de référence

```
# contrasts(mort$age)
contrasts((mort$periode))
```

```

## 48-52 53-57 58-62 63-67 68-72 73-77 78-82 83-87 88-92
## 43-47 0 0 0 0 0 0 0 0 0
## 48-52 1 0 0 0 0 0 0 0 0
## 53-57 0 1 0 0 0 0 0 0 0
## 58-62 0 0 1 0 0 0 0 0 0
## 63-67 0 0 0 1 0 0 0 0 0
## 68-72 0 0 0 0 1 0 0 0 0
## 73-77 0 0 0 0 0 1 0 0 0
## 78-82 0 0 0 0 0 0 1 0 0
## 83-87 0 0 0 0 0 0 0 1 0
## 88-92 0 0 0 0 0 0 0 0 1

```

```
contrasts(mort$sex)
```

```
##     F  
## H 0  
## F 1
```

La modalité de référence est l'âge zéro. La modalité de référence pour la période temporelle est entre 1943 et 1947. Enfin la modalité de référence au niveau sexe est l'homme. **TODO changer la modalité de ref de l'âge**

TODO choisir un modèle

```
# glm(nb.deces ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Infections ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Cancer ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Tumeurs.Bénignes ~ periode * sex * age, offset = log(PA), data = mort)  
endocrine <- glm(Endocrine ~ periode + sex + age, offset = log(PA),  
                  data = mort)  
# glm(Maladies.Sang ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Cerebrovasculaire ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Cardiaque ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Maladies.Respiratoires ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Maladies.Foie ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Maladies.Digestives ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Maladie.Génito_urinaire ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Non.Definis ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Mort.Naturelle ~ periode * sex * age, offset = log(PA), data = mort)  
# glm(Mort.Violente ~ periode * sex * age, offset = log(PA), data = mort)
```